



Deep Latent Factor Model for Predicting Drug Target Interactions

Aanchal Mongia, Vidit Jain, Emilie Chouzenoux, Angshul Majumdar

► To cite this version:

Aanchal Mongia, Vidit Jain, Emilie Chouzenoux, Angshul Majumdar. Deep Latent Factor Model for Predicting Drug Target Interactions. ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing, May 2019, Brighton, United Kingdom. 10.1109/icassp.2019.8683123 . hal-02406989

HAL Id: hal-02406989

<https://hal.science/hal-02406989>

Submitted on 12 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deep Latent Factor Model for Predicting Drug Target Interactions

Aanchal Mongia
Dept. of CS, IIIT-Delhi
aanchalm@iiitd.ac.in

Vidit Jain
Dept. of CS, IIIT-Delhi
vidit17121@iiitd.ac.in

Emilie Chouzenoux
LIGM, Univ. Paris East
CVN, CentraleSupélec
chouzeno@univ-mlv.fr

Angshul Majumdar
Dept. of ECE, IIIT-Delhi
angshul@iiitd.ac.in

Abstract— In drug target interaction (DTI) the interactions of some (a subset) drugs on some (a subset) targets are known. The goal is to predict the interactions of all drugs on all targets. One approach is to formulate this as a matrix completion problem, where the matrix of interactions having drugs along the rows and targets along the columns is partially filled. So far standard matrix completion approaches such as nuclear norm minimization and matrix factorization have been used to address the problem. In this work, we propose a deep matrix factorization approach to improve the prediction results. Experiments have been performed on benchmark databases and comparison carried out with some state-of-the-art algorithms. Empirically our proposed deep method, outperforms all the techniques compared against.

Keywords— *matrix completion, deep learning, matrix factorization, drug target interaction*

I. INTRODUCTION

The traditional approach to verify interactions between drugs and target proteins relies on wet lab experiments. However, such experiments are costly, laborious and time consuming. Therefore, there is a need to complement wet lab experiments by computational means. Wet lab experiments still remain the gold standard, but by computational techniques one can narrow down the search space for the experiments to be conducted in such labs.

An important application of predicting drug target interactions (DTI) is drug repositioning. Repositioning means reuse of an existing drug developed to cure a certain disease for treating a new / different disease. Obviously, in theory, the interactions can be found in wet labs, but the cost and time taken to find out all the targets for every drug would be humongous. Therefore, there is a need to use efficient computational techniques to reduce costs and accelerate the drug discovery process.

One classic example of efficient repositioning is Gleevec (imatinib mesylate); this drug was originally thought to interact only with the Bcr-Abl fusion gene associated with leukemia. It was later discovered that Gleevec also interacts with PDGF and KIT, eventually leading it to be repositioned to treat gastrointestinal stromal tumors as well [1, 2]. A drug's ability to interact with multiple targets may contribute to its polypharmacology (i.e. having multiple therapeutic effects),

which is a clear motivation to apply computational DTIs on existing drugs.

In the past, many approaches have been proposed towards computational DTI. We will briefly discuss them. This will help us in positioning our proposed work.

At the onset, we discuss the similarity between the DTI problem and the problem of collaborative filtering (CF). CF is a standard problem in information retrieval. It is used in recommendations systems (e.g. in Netflix movie recommendations and Amazon product recommendations). It relies on a database of user's and their ratings on items (movies, products, etc.). Obviously, not all the ratings are available; users typically rate only a small subset of items. The objective is thus to estimate the missing ratings, for all the users on all the items. If that can be done accurately, recommendation accuracy increases. The similarity between DTI and CF should be straightforward now; the drugs play the role of users and the targets play the role of items. The interactions are similar to the ratings. Over the years, many approaches originally developed for CF have been leveraged to solve the DTI problems.

In both CF [3] and DTI [4, 5] the initial techniques were based on simple neighborhood-based models. To predict the interaction of a (active) drug on a target, the first step is to find out similar (neighbor) drugs by computing some kind of a similarity score. Once the neighborhood is obtained, the interaction values for the drugs in the neighborhood are weighted (by the normalized similarity score) to interpolate the interaction of the active drug on the target.

The second approach was based on bipartite local models. In such models, a local model is built for every drug and target. For example, in [6], an SVM was trained to predict interaction of each drug on all targets and each target on all drugs. Finally, the decision from the two were fused. This is just an example, there are other techniques falling under this generic approach like [7, 8].

The third category is based on network diffusion models. One technique for DTI prediction based on such models relies on a random walk strategy on the network with a predefined transition matrix [9]. Another work falling under this category, predicts interactions by finding simple (without loops) connections between nodes of the network.

The fourth approach is based on matrix factorization. These techniques were originally developed for collaborative filtering [11]. It is assumed that the drugs and targets are characterized by latent factors. The probability of interaction is high when the latent factors match; i.e. when the inner product has a high value. Therefore, it is logical to express the interaction matrix as a (inner) product of drug and target latent factors. This allows matrix factorization (and its variants) to be applied [12, 13].

The fifth and final approach is based on classification. The chemical / biological information is used to generate features for drugs and targets individually. The two features are then concatenated and the corresponding interaction is assumed to the class corresponding to this feature. Any standard classifier can be used for the final classification. In such class of techniques, the emphasis is put on different feature selection mechanisms [14, 15].

In a very recent review paper [16] it was empirically shown that matrix factorization-based techniques yield by far the best results. Using this statement as an initial point, we propose to develop a deep technique for matrix factorization. In standard (shallow) matrix factorization, the data matrix is decomposed / factored into two factors / matrices. We extend this by allowing factorization into more than two; thus, making it deep. The ensuing problem is solved using the proximal alternating projections well-suited to tackle non-convex problems with block structure.

The relevant literature has already been reviewed in section I. In the following section II, the proposed approach will be discussed. The experimental results will be described in section III. Conclusions of this work and further extensions will be discussed in section IV.

II. PROPOSED APPROACH

A. Problem Formulation and Related Techniques

Let us organize the drug target interactions as a matrix X . Along the rows are the drugs and along the columns are the targets. The entries in the matrix are the interactions. This matrix is partially observed, because not all drug target interactions are known. This is expressed as,

$$Y = R \cdot X \quad (1)$$

Here R is a binary sampling matrix having 1's in positions where the interaction is known and 0's elsewhere; Y is the available partially sampled DTI matrix.

The objective is to estimate the matrix X given Y and knowing R . In factorization-based techniques such as [12, 13], it is assumed that the matrix X is low rank, i.e. only a few factors are responsible for drugs interacting with targets. Explaining the biological significance is beyond the scope of this paper; the interested readers should peruse the aforesaid references. Assuming that X is low rank, one can factor it into two matrices U and V .

$$X_{M \times N} = U_{M \times r} V_{r \times N}, r \ll (m, n) \quad (2)$$

Here M and N are the number of drugs and targets respective and r is the presumed rank of the matrix.

The standard approach to solve (2) requires some regularized matrix factorization technique. In general, this can be written as follows,

$$\min_{U, V} \|Y - R \cdot (UV)\|_F^2 + \lambda \rho(U) + \lambda \rho(V) \quad (3)$$

Here ρ denotes some regularization function. It can be as simple as a Tikhonov penalty or can be sophisticated like graph Laplacian.

The first term of the cost function involved in (3) is bilinear and hence non-convex. An alternate convex formulation has been proposed to the matrix completion problem, which consists in directly recovering the matrix X by minimizing its nuclear norm [17-19].

$$\min_R \|Y - R \cdot X\|_F^2 + \lambda \|X\|_{NN} \quad (4)$$

The nuclear norm $\|X\|_{NN}$ is defined as the sum of singular values of the matrix. It is the nearest convex surrogate to the rank of a matrix. Exact rank minimization is supposed to be NP hard.

Although mathematically sound, the nuclear norm minimization technique has never been used for DTI prediction. This is largely because solving (6) is time consuming; and it is not easy to incorporate biological / chemical metadata into the formulation. Researchers stick to matrix factorization for predicting DTI.

Some recent studies [20-22] proposed factorizing the data matrix as a product of more than two factors. This is expressed by,

$$X = U_1 U_2 \dots U_N V \quad (5)$$

Without any additional constraint and/or non-linearity, formulation (5) is bound to collapse to shallow matrix factorization. This is prevented by enforcing positivity constraint on the factors during optimization. This acts as a rectified linear unit (ReLU) type activation.

Note that the aforementioned works [20-22] are not directly relevant to us as they propose a solution when the data matrix is fully observed. This is not the case for drug target interactions.

B. Deep Latent Factor Model

The term latent factor stems from collaborative filtering area. There in, the matrices U and V are supposed to represent hidden variables. The concept can be applied to DTI as well. This is the reason, we name our method deep latent factor model to differentiate from earlier studies [20-22] which have been named 'deep matrix factorization'.

Following the deep matrix factorization approach, we propose to factor the DTI matrix X into multiple layers of latent factors / matrices. The complete model is given by,

$$Y = R \cdot X = R \cdot (U_1 U_2 \dots U_N V) \quad (6)$$

We will incorporate ReLU type activation during the optimization process in order to enforce non-negativity over each variable.

We will explicitly derive our algorithm in the case of three layers; note that it is generic enough for more or fewer layers. The objective function to solve (for three layers) is :

$$\min_{U_1, U_2, U_3, V} \frac{1}{2} \|Y - R \cdot (U_1 U_2 U_3 V)\|_F^2 \text{ such that} \quad (7)$$

$$U_1 U_2 U_3 V \geq 0, U_1 \geq 0, U_2 \geq 0, U_3 \geq 0, V \geq 0$$

This is equivalent to solve,

$$\min_{U_1, U_2, U_3, V, X} \frac{1}{2} \|Y - R \cdot (U_1 U_2 U_3 V)\|_F^2 \text{ such that} \quad (8)$$

$$U_1 U_2 U_3 V = X \text{ and } X \geq 0, U_1 \geq 0, U_2 \geq 0, U_3 \geq 0, V \geq 0$$

The Lagrangian-based formulation reads, with $\lambda > 0$,

$$\min_{U_1, U_2, U_3, V, X} \frac{1}{2} \|Y - R \cdot (U_1 U_2 U_3 V)\|_F^2 + \frac{\lambda}{2} \|U_1 U_2 U_3 V - X\|_F^2 \quad (9)$$

$$\text{such that } X \geq 0, U_1 \geq 0, U_2 \geq 0, U_3 \geq 0, V \geq 0$$

Problem (9) fits within the framework studied in [23, 38] regarding proximal alternating schemes for non-convex optimization. Let us introduce the notation:

$$G(X, U_1, U_2, U_3, V) = \frac{1}{2} \|Y - R \cdot X\|_F^2 + \frac{\lambda}{2} \|U_1 U_2 U_3 V - X\|_F^2$$

Then the minimization algorithm reads, with the parameters

$$(\theta_1, \theta_2, \theta_3, \theta) > 0 \text{ and } \gamma \leq (\|R\|_F^2 + \lambda)^{-1} :$$

Initialize: $X^0, U_1^0, U_2^0, U_3^0, V^0$

For $k = 1, 2, \dots$

$$X^{k+1} = P_+ \left(X^k - \gamma (R^T \cdot (R \cdot X^k - Y) + \lambda (X^k - U_1^k U_2^k U_3^k V^k)) \right)$$

$$U_1^{k+1} = P_+ \left(U_1^k - \theta_1 \lambda (U_1^k U_2^k U_3^k V^k - X^{k+1}) (U_2^k U_3^k V^k)^T \right)$$

$$U_2^{k+1} = P_+ \left(U_2^k - \theta_2 \lambda (U_1^{k+1})^T (U_1^{k+1} U_2^k U_3^k V^k - X^{k+1}) (U_3^k V^k)^T \right)$$

$$U_3^{k+1} = P_+ \left(U_3^k - \theta_3 \lambda (U_1^{k+1} U_2^{k+1})^T (U_1^{k+1} U_2^{k+1} U_3^k V^k - X^{k+1}) (V^k)^T \right)$$

$$V^{k+1} = \text{prox}_{\theta G(X^{k+1}, U_1^{k+1}, U_2^{k+1}, U_3^{k+1}, \cdot)}(V^k)$$

Hereabove, P_+ denotes the simple projection on the positive orthant. Moreover, the last step of our algorithm computes the proximity operator [36] of function θG , with respect to his last variable V . This update has an explicit expression. Indeed, by definition,

Let $\theta > 0$

$$\hat{V} = \text{prox}_{\theta G(X, U_1, U_2, U_3, \cdot)}(\bar{V})$$

$$= \arg \min_V \theta G(X, U_1, U_2, U_3, V) + \frac{1}{2} \|V - \bar{V}\|_F^2$$

$$= \arg \min_V \frac{\theta \lambda}{2} \|U_1 U_2 U_3 V - X\|_F^2 + \frac{1}{2} \|V - \bar{V}\|_F^2$$

In a vectorized form (stacking all matrix columns vertically), the solution of the above problem is:

$$\hat{v} = (\lambda \theta B^T B + I)^{-1} (\lambda \theta B^T y + \bar{v}) \quad (10)$$

with $B = U_1 U_2 U_3 \otimes I$, where I denotes the identity matrix, and \otimes is the Kronecker product [37]. \hat{V} can then be obtained by reshaping from vector to matrix form.

The convergence of the iterates generated by our method, to a critical point of problem (9) is guaranteed, using the analysis from [23, 38].

III. EXPERIMENTAL RESULTS

We use the four benchmark datasets introduced in [24] concerning four different classes of target proteins, namely enzymes (Es), ion channels (ICs), G protein-coupled receptors (GPCRs) and nuclear receptors (NRs). The data was simulated from public databases KEGG BRITE [25], BRENDA [26] SuperTarget [27] and DrugBank [28]. Table I summarizes the statistics of the four datasets.

TABLE I: DATASET DESCRIPTION

Datasets	NR	GPCR	IC	E
Drugs	54	223	201	445
Targets	26	95	204	664
Interactions	90	635	1476	2926

The data gathered from these databases is formatted as an adjacency matrix, called interaction matrix between drugs and targets, encoding the interaction between n drugs and m targets as 1 if the drug d^i and target t^j are known to interact and 0, otherwise.

It is customary in this area to evaluate algorithms based on two well-known metrics AUC (area under the curve) and AUPR (area under the precision-recall curve).

A receiver operating characteristics (ROC) is simply a plot showing the true positive rate for a method as a function of the false positive rate. AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. The higher it is, the better the model is.

AUPR punishes highly ranked false positives much more than AUC, this point being important practically since only highly ranked drug-target pairs in prediction will be biologically or chemically tested later in a usual drug discovery process, meaning that highly ranked false positives should be avoided. The precision-recall curve shows the tradeoff between precision and recall for different threshold. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).

We have compared our technique with standard matrix completion techniques. For nuclear norm minimization, we used the popular singular value thresholding (SVT) [29] algorithm. Since the drug target interactions forms a binary matrix, we also tried the one bit matrix completion (BMC) algorithm [30]. As the benchmark for matrix factorization we have used the probabilistic matrix factorization (PMF) [31]. The standard (shallow) matrix factorization formulation is equivalent to our formulation in the one layer case; so we have used our proposed method also for this case. We have also compared with the state-of-the-art in DTI [39] –graph regularized matrix factorization (GRMF).

TABLE II: LATENT FACTORS – INPUTS TO ALGORITHMS

Layers	E	GPCR	IC	NR
1 Layer / SVT / PMF / BMC	10	80	85	25
2 Layer	185-25	40-10	100-15	20-10
3 Layer	180-85-15	65-15-5	100-50-10	20-10-5

The dataset was divided randomly into 70% training set and rest 30% testing set. The splitting was done 10 times and the average results are reported. For all the techniques we tuned the parameters by cross-validation on the training data. All the approaches had only one regularization parameter to tune. But they require specifying the rank / number of latent factors. These values are given in Table II.

TABLE III: TABLE SHOWING AUPR

Dataset	1 Layer	2 Layer	3 Layer	SVT	BMC	PMF	GRMF
E	.639	.714	.728	.010	.706	.622	.498
GPCR	.599	.615	.616	.036	.604	.556	.442
IC	.792	.828	.828	.056	.803	.760	.381
NR	.097	.121	.125	.092	.107	.091	.097

TABLE IV: TABLE SHOWING AUC

Dataset	1 Layer	2 Layer	3 Layer	SVT	BMC	PMF	GRMF
E	.877	.897	.899	.496	.879	.855	.573
GPCR	.870	.881	.884	.528	.876	.858	.561
IC	.928	.942	.941	.488	.931	.874	.628
NR	.634	.669	.669	.461	.639	.618	.456

Note that the results cannot be compared with the published work [39]. This is because, GRMF boosts the results by pre-processing. Since our objective is to compare different algorithms for DTI, we have refrained from using any pre-processing. Without pre-processing, GRMF produces poor results as can be seen here.

The detailed results are shown in Tables III and IV. We see that our method consistently outperforms the others. For some cases the two layers formulation yields the best results, but for most the three layer decomposition performs the best. The reason, we cannot improve too much by going deeper is because of the size of the datasets. They are small, and hence going deeper does not help. In one hand going deeper, improves abstraction of the latent factors but on the other, the number of parameters to learn also increases. Owing to the second fact, over-fitting sets in, the algorithm fails to generalize on the unseen data and hence the results

deteriorate. In fact, we tried going to 4 layers, but the results were worse than for a single layer.

The single layer formulation is the basic matrix factorization method and performs the worst. One bit matrix completion is tailor-made for the problem and hence performs reasonably well – better than any other shallow technique. The singular value thresholding based nuclear norm minimization technique yields the worst results. Probabilistic matrix factorization also yields decent results, but is poorer than ours and one bit matrix completion.

However, one must note that one bit matrix completion will not be of much use once biologically inspired pre-processing steps such as [39] are in place to boost the results. These steps will convert the binary matrix into real numbers and hence BMC will not be applicable.

IV. CONCLUSION

This work introduces the deep latent factor model for matrix completion. So far matrix completion was solved by shallow techniques. Either the partially sampled matrix was directly recovered by exploiting its rank deficiency via nuclear norm minimization; else the matrix was decomposed into two factors and recovered via matrix factorization. This is the first work that shows, how the matrix can be decomposed into more than two factors and estimated via the product of the individual factors.

We have applied our proposed technique to solve predicting drug target interactions. We have shown that when the associated biological metadata of the drugs or the targets is not available, our method yields the best possible results (compared to shallow matrix completion, nuclear norm minimization and autoencoder based imputation).

We propose to extend the work in two ways. First, our algorithm is generic. It can be applied to other matrix completion problems like collaborative filtering, dynamic MRI reconstruction [32, 33], single cell RNA sequence imputation [34, 35] etc. It remains to be seen, if the improvements shown here can be reproduced for other applications as well.

The second extension of this work would be to better understand the biological significance of our problem and modify our algorithm to exploit biological metadata associated with the drugs and targets. In past works such [12, 13], it has been shown that incorporating such metadata improves the predictions results by a large extent. Also, most we would like to see how different pre-processing mechanisms such as [39] help improve the results even further.

ACKNOWLEDGEMENT

This work is supported by Indo-French Grant DST-CNRS-2016-02 NextGenBP.

REFERENCES

- [1] Frantz, Simon. "Drug discovery: playing dirty." *Nature* (2005): 942.
- [2] McLean, Sean R., Mali Gana-Weisz, Basil Hartzoulakis, Richard Frow, Jeremy Whelan, David Selwood, and Chris Boshoff. "Imatinib binding and cKIT inhibition is abrogated by the cKIT kinase domain I

- missense mutation Val654Ala." *Molecular cancer therapeutics* 4, no. 12 (2005): 2008-2015.
- [3] Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl. "Item-based collaborative filtering recommendation algorithms." In *Proceedings of the 10th international conference on World Wide Web*, pp. 285-295. ACM, 2001.
 - [4] Yamanishi, Yoshihiro, Michihiro Araki, Alex Gutteridge, Wataru Honda, and Minoru Kanehisa. "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces." *Bioinformatics* 24, no. 13 (2008): i232-i240.
 - [5] Shi, J.Y. and Yiu, S.M., 2015, November. SRP: A concise non-parametric similarity-rank-based model for predicting drug-target interactions. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1636-1641). IEEE.
 - [6] Bleakley, Kevin, and Yoshihiro Yamanishi. "Supervised prediction of drug-target interactions using bipartite local models." *Bioinformatics* 25, no. 18 (2009): 2397-2403.
 - [7] van Laarhoven, Twan, Sander B. Nabuurs, and Elena Marchiori. "Gaussian interaction profile kernels for predicting drug-target interaction." *Bioinformatics* 27, no. 21 (2011): 3036-3043.
 - [8] Mei, J.P., Kwok, C.K., Yang, P., Li, X.L. and Zheng, J., 2012. Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics*, 29(2), pp.238-245.
 - [9] Chen, Xing, Ming-Xi Liu, and Gui-Ying Yan. "Drug-target interaction prediction by random walk on the heterogeneous network." *Molecular BioSystems* 8, no. 7 (2012): 1970-1978.
 - [10] Ba-Alawi, Wail, Othman Soufan, Magbubah Essack, Panos Kalnis, and Vladimir B. Bajic. "DASPfind: new efficient method to predict drug-target interactions." *Journal of cheminformatics* 8, no. 1 (2016): 15.
 - [11] Koren, Y., Bell, R. and Volinsky, C., 2009. Matrix factorization techniques for recommender systems. *Computer*, (8), pp.30-37.
 - [12] Cobanoglu, Murat Can, Chang Liu, Feizhuo Hu, Zoltán N. Oltvai, and Ivet Bahar. "Predicting drug-target interactions using probabilistic matrix factorization." *Journal of chemical information and modeling* 53, no. 12 (2013): 3399-3409.
 - [13] Ezzat, Ali, Peilin Zhao, Min Wu, Xiao-Li Li, and Chee-Keong Khoo. "Drug-target interaction prediction with graph regularized matrix factorization." *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 14, no. 3 (2017): 646-656.
 - [14] He, Zhisong, Jian Zhang, Xiao-He Shi, Le-Le Hu, Xiangyin Kong, Yu-Dong Cai, and Kuo-Chen Chou. "Predicting drug-target interaction networks based on functional groups and biological features." *PloS one* 5, no. 3 (2010): e9603.
 - [15] Yu, Hua, Jianxin Chen, Xue Xu, Yan Li, Huihui Zhao, Yupeng Fang, Xiuxiu Li, Wei Zhou, Wei Wang, and Yonghua Wang. "A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data." *PloS one* 7, no. 5 (2012): e37608.
 - [16] Ezzat, Ali, Min Wu, Xiao-Li Li, and Chee-Keong Khoo. "Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey." *Briefings in bioinformatics* (2018).
 - [17] Candès, Emmanuel J., and Terence Tao. "The power of convex relaxation: Near-optimal matrix completion." *IEEE Transactions on Information Theory* 56, no. 5 (2010): 2053-2080.
 - [18] Candès, Emmanuel J., and Benjamin Recht. "Exact matrix completion via convex optimization." *Foundations of Computational mathematics* 9, no. 6 (2009): 717.
 - [19] Recht, Benjamin. "A simpler approach to matrix completion." *Journal of Machine Learning Research*, no. 12, pp. 3413-3430, 2011.
 - [20] Trigeorgis, George, Konstantinos Bousmalis, Stefanos Zafeiriou, and Björn W. Schuller. "A deep matrix factorization method for learning attribute representations." *IEEE transactions on pattern analysis and machine intelligence* 39, no. 3 (2017): 417-429.
 - [21] Li, Zechao, and Jinhui Tang. "Weakly supervised deep matrix factorization for social image understanding." *IEEE Transactions on Image Processing* 26, no. 1 (2017): 276-288.
 - [22] Trigeorgis, George, Konstantinos Bousmalis, Stefanos Zafeiriou, and Bjoern Schuller. "A deep semi-nmf model for learning hidden representations." In *International Conference on Machine Learning*, pp. 1692-1700. 2014.
 - [23] Chouzenoux Emilie, Pesquet Jean-Christophe and Repetti Audrey. A Block Coordinate Variable Metric Forward-Backward Algorithm. *Journal of Global Optimization*, Vol. 66, No. 3, pages 457–485, 2016.
 - [24] Yamanishi, Yoshihiro, Michihiro Araki, Alex Gutteridge, Wataru Honda, and Minoru Kanehisa. "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces." *Bioinformatics* 24, no. 13 (2008): i232-i240.
 - [25] Kanehisa, Minoru, Susumu Goto, Masahiro Hattori, Kiyoko F. Aoki-Kinoshita, Masumi Itoh, Shuichi Kawashima, Toshiaki Katayama, Michihiro Araki, and Mika Hirakawa. "From genomics to chemical genomics: new developments in KEGG." *Nucleic acids research* 34, no. suppl_1 (2006): D354-D357.
 - [26] Schomburg, Ida, Antje Chang, Christian Ebeling, Marion Gremse, Christian Heldt, Gregor Huhn, and Dietmar Schomburg. "BRENDA, the enzyme database: updates and major new developments." *Nucleic acids research* 32, no. suppl_1 (2004): D431-D433.
 - [27] Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Urdiales, E.G., Gewiss, A., Jensen, L.J. and Schneider, R., 2007. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic acids research*, 36(suppl_1), pp.D919-D922.
 - [28] Wishart, David S., Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. "DrugBank: a knowledgebase for drugs, drug actions and drug targets." *Nucleic acids research* 36, no. suppl_1 (2007): D901-D906.
 - [29] Cai, Jian-Feng, Emmanuel J. Candès, and Zuowei Shen. "A singular value thresholding algorithm for matrix completion." *SIAM Journal on Optimization* 20, no. 4 (2010): 1956-1982.
 - [30] Davenport, Mark A., Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. "1-bit matrix completion." *Information and Inference: A Journal of the IMA* 3, no. 3 (2014): 189-223.
 - [31] Mnih, Andriy, and Ruslan R. Salakhutdinov. "Probabilistic matrix factorization." In *Advances in neural information processing systems*, pp. 1257-1264. 2008.
 - [32] Zhao, Bo, Justin P. Haldar, Cornelius Brinegar, and Zhi-Pei Liang. "Low rank matrix recovery for real-time cardiac MRI." In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pp. 996-999. IEEE, 2010.
 - [33] Trzasko, J., A. Manduca, and E. Borisch. "Local versus global low-rank promotion in dynamic MRI series reconstruction." In *Proc. Int. Symp. Magn. Reson. Med*, p. 4371. 2011.
 - [34] Gong, Wuming, Il-Youp Kwak, Pruthvi Pota, Naoko Koyano-Nakagawa, and Daniel J. Garry. "DrImpute: imputing dropout events in single cell RNA sequencing data." *BMC bioinformatics* 19, no. 1 (2018): 220.
 - [35] Huang, Mo, Jingshu Wang, Eduardo Torre, Hannah Dueck, Sydney Shaffer, Roberto Bonasio, John I. Murray, Arjun Raj, Mingyao Li, and Nancy R. Zhang. "SAVER: gene expression recovery for single-cell RNA sequencing." *Nature Methods* 15, no. 7 (2018): 539.
 - [36] Combettes Patrick-Louis and Pesquet Jean-Christophe. "Proximal Splitting Methods in Signal Processing". In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.
 - [37] Van Loan Charles. "The Ubiquitous Kronecker Product", *Journal of Computational and Applied Mathematics*, Vol. 123, no. 1-2, pp. 85-100, 2000.
 - [38] Attouch Hedi, Bolte Jérôme, Redont Patrick and Soubeyran Antoine. "Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality". *Math. Oper. Res.* 35(2), 438–457 (2010).
 - [39] Ezzat, Ali, Peilin Zhao, Min Wu, Xiao-Li Li, and Chee-Keong Khoo. "Drug-target interaction prediction with graph regularized matrix factorization." *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 14, no. 3 (2017): 646-656.